

PRINCIPAL COMPONENT ANALYSIS AS A TOOL FOR ENHANCED WELL LOG INTERPRETATION

BOGDAN MIHAI NICULESCU, GINA ANDREI

*University of Bucharest, Faculty of Geology and Geophysics, Department of Geophysics,
6, Traian Vuia St., 020956 Bucharest, Romania
(bogdan.niculescu@gg.unibuc.ro; bogdan.mihai.niculescu@gmail.com; gina.andrei@gg.unibuc.ro)*

We investigate the potential usefulness of Principal Component Analysis (PCA) method in providing meaningful petrophysical information, in addition to the results obtained via conventional well log interpretation, or to constrain and validate such results. We applied PCA to a geophysical logging data set recorded in a natural gas exploration well drilled in the NW part of Moldavian Platform – Romania. The first principal components of the data seem to respond to major lithological changes or shale/clay content variations, whereas the higher-order principal components most likely reflect fluid-related data variability, such as fluids type and/or volume. The results of this study suggest that PCA may successfully complement the standard log interpretation and formation evaluation methods.

Key words: Principal Component Analysis, Moldavian Platform (Romania), natural gas, geophysical well logs, log interpretation.

1. INTRODUCTION

Principal Component Analysis (PCA) (Pearson, 1901; Hotelling, 1933; Jolliffe, 2002) is a multivariate data dimensionality reduction technique, used to simplify a data set to a smaller number of factors that explain most of the variability (variance). PCA aims to convert a set of correlated variables to a number of uncorrelated orthogonal *principal components* (PCs). Besides dimensionality reduction, this analysis may also be employed to discover and interpret the dependencies and relationships possibly existing among the original variables. PCA is a linear transformation that maps the data in a new (rotated) coordinate system, such that the new variables are linear combinations of the original variables and they summarize the dominant data trends. In practice, PCA is carried out by computing the covariance matrix of the data set, and then the eigenvalues and eigenvectors of the covariance matrix are computed and sorted according to decreasing eigenvalues, *i.e.* decreasing amounts of data variability. For a meaningful interpretation of the principal components it is important to determine which original variables are associated with particular components. PCA's component sorting based on the amount of variance criterion is not always relevant or

significant; features with low variance may actually have high predictive relevance and importance, depending upon the application.

PCA has been successfully used for a variety of well logging data applications, such as: identification and characterization of pressure seals / low permeability intervals (Moline *et al.*, 1992), delineation of lithostratigraphic units, identification of aquifer formations and distinction between hydraulic flow units (Kassenaar, 1991; Barrash, Morin, 1997; Gonçalves, 1998), interdependency and correlation between some hydraulic properties and geophysical / petrophysical parameters (Morin, 2006), well-to-well correlation by pattern recognition (Lim *et al.*, 1998) etc. In this study we investigate and discuss the potential usefulness of PCA in providing meaningful petrophysical information in the case of hydrocarbon exploration wells, in addition to the results obtained via conventional log interpretation, or in order to constrain and validate such results.

2. SUMMARY OF PRINCIPAL COMPONENT ANALYSIS METHOD

Taking into account a multivariate data set X consisting in p random variables $x_1, x_2, \dots, x_i, \dots$,

x_p (i.e., geophysical well logs, each log consisting in n measurements of a specific subsurface property), the p principal components $z_1, z_2, \dots, z_i, \dots, z_p$ of the data set (alternate notation: $PC_1, PC_2, \dots, PC_i, \dots, PC_p$) are given by the linear combinations

$$z_i = \mathbf{a}_i^T \mathbf{X} = a_{i1} x_1 + a_{i2} x_2 + \dots + a_{ip} x_p; i = 1, 2, \dots, p \quad (1)$$

where \mathbf{a}_i are the column vectors of an orthogonal p -by- p transformation matrix \mathbf{A} ($\mathbf{A}^T \mathbf{A} = \mathbf{A} \mathbf{A}^T = \mathbf{I}$, with T denoting the transpose and \mathbf{I} representing the p -by- p identity matrix). Besides a normalization condition expressed by $\mathbf{a}_i^T \mathbf{a}_i = 1$ ($i = 1, 2, \dots, p$) and the orthogonality of the PCs, a condition imposed when extracting the PCs is $\text{var}(z_1) \geq \text{var}(z_2) \geq \dots \geq \text{var}(z_p)$, where var stands for the variance. The first PC is $\mathbf{a}_1^T \mathbf{X}$, subject to $\mathbf{a}_1^T \mathbf{a}_1 = 1$, that maximizes $\text{var}(\mathbf{a}_1^T \mathbf{X})$; the second PC is $\mathbf{a}_2^T \mathbf{X}$ that maximizes $\text{var}(\mathbf{a}_2^T \mathbf{X})$, subject to $\mathbf{a}_2^T \mathbf{a}_2 = 1$ and covariance $\text{cov}(\mathbf{a}_1^T \mathbf{X}, \mathbf{a}_2^T \mathbf{X}) = 0$ (uncorrelated principal components) and so on. Generally, the i -th PC $z_i = \mathbf{a}_i^T \mathbf{X}$, subject to $\mathbf{a}_i^T \mathbf{a}_i = 1$, maximizes $\text{var}(\mathbf{a}_i^T \mathbf{X})$ with $\text{cov}(\mathbf{a}_i^T \mathbf{X}, \mathbf{a}_k^T \mathbf{X}) = 0$, for $k < i$.

For each PC, the variance that has to be maximized subject to the condition $\mathbf{a}_i^T \mathbf{a}_i = 1$ (i.e., $\mathbf{a}_i^T \mathbf{a}_i - 1 = 0$) can be expressed as

$$\text{var}(z_i) = \text{var}(\mathbf{a}_i^T \mathbf{X}) = \mathbf{a}_i^T \boldsymbol{\Sigma} \mathbf{a}_i \rightarrow \text{maximum}, \quad (2)$$

where $\boldsymbol{\Sigma}$ is the p -by- p sample covariance matrix of the data set. The constrained maximization problem can be solved by creating a function

$$L = \mathbf{a}_i^T \boldsymbol{\Sigma} \mathbf{a}_i - \lambda (\mathbf{a}_i^T \mathbf{a}_i - 1), \quad (3)$$

where λ stands for a Lagrange multiplier. By cancelling the partial derivatives of function L with respect to the unknown \mathbf{a}_i vectors, i.e. $\partial L / \partial \mathbf{a}_i = 0$, one obtains the matrix equation

$$(\boldsymbol{\Sigma} - \lambda \mathbf{I}) \mathbf{a}_i = 0. \quad (4)$$

The characteristic equation $\det(\boldsymbol{\Sigma} - \lambda \mathbf{I}) = 0$ has p roots (eigenvalues) λ_i , $i = 1, 2, \dots, p$, such that $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$. Once the eigenvalues λ_i are determined, the corresponding eigenvectors \mathbf{a}_i can be computed by solving Eq. (4). For a p

variables data set \mathbf{X} , each \mathbf{a}_i is a p -by-1 vector defining the axes of a new, rotated coordinates system that maximizes data variability along each axis (Fig. 1). PCA's results are usually expressed and interpreted in terms of *component scores* (z_i values corresponding to particular data points) and *loadings* (the components of each eigenvector \mathbf{a}_i , i.e. $a_{i1}, a_{i2}, \dots, a_{ip}$ from Eq. (1), which act as weighting factors of the original variables $x_1, x_2, \dots, x_i, \dots, x_p$).

Software implementations of PCA are available as dedicated modules within well log interpretation packages (e.g., the "Principal Component Analysis" module from *Interactive Petrophysics (IP™)* software, © LR Senergy). In the *MATLAB™* (© MathWorks) programming environment PCA can be carried out by using the built-in functions **corrcoef**, **zscore**, **cov** and **pcacov** in a code such as

```
clear all; close all; clc
load DataMatrix
CorrelationMatrix = corrcoef(DataMatrix)
Data = zscore(DataMatrix);
CovarianceMatrix = cov(Data)
[COEFF, latent, explained] = pcacov
                                (CovarianceMatrix)
SCORE = Data*COEFF;
save 'COEFF.txt' COEFF -ascii
save 'LATENT.txt' latent -ascii
save 'PERCENT.txt' explained -ascii
save 'SCORE.txt' SCORE -ascii
```

where: *DataMatrix* = n -by- p matrix \mathbf{X} storing p geophysical well logs with n samples/log; *COEFF* = p -by- p matrix storing the PC coefficients (the loadings \mathbf{a}_i); *latent* = vector storing the PC variances (eigenvalues λ_i of the covariance matrix); *SCORE* = the computed linear combinations $z_i = \mathbf{a}_i^T \mathbf{X}$ for each depth level.

Figure 1 illustrates the principle of PCA method, taking into account the case of two random variables x_1 and x_2 .

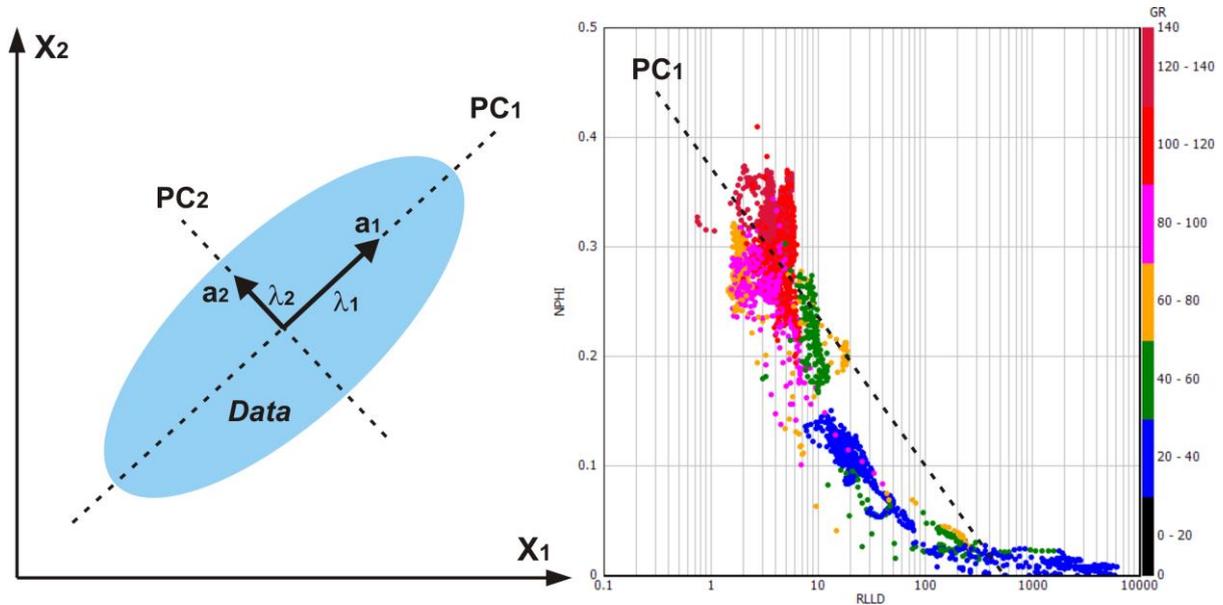


Fig. 1 – Left: Idealized illustration of the PCA method for the case of two random variables x_1 and x_2 . PCA finds the main variability directions in the data "cloud" and defines a new coordinate system, using optimal rotations. The axes of this system are defined by the eigenvectors a_1 and a_2 . The eigenvalues λ_1 and λ_2 ($\lambda_1 \geq \lambda_2$) correspond to the data variance in the newly defined coordinate system. Right: Interdependency between two real random variables (geophysical logs recorded in the exploration well analyzed in this paper – apparent neutron porosity Φ_N vs. deep resistivity ρ_{LLD}). The main variability direction shown corresponds to the first principal component (PC_1).

3. APPLICATION OF PRINCIPAL COMPONENT ANALYSIS METHOD ON A BOREHOLE GEOPHYSICAL DATA SET (GAS EXPLORATION WELL, MOLDAVIAN PLATFORM – ROMANIA)

In order to study the applicability and effectiveness of the PCA method, we have processed and interpreted a wireline logging data set from a gas (biogenic methane) exploration well drilled in the Moldavian Platform – Romania. The PCA results were evaluated by comparison with the results of conventional log interpretation and with additional information (production tests, lithology logs and actual formation tops).

3.1. GEOLOGICAL AND TECTONIC SETTING

The Moldavian Platform, located in the NE part of Romania, is the oldest platform unit of the Romanian territory and represents the SW termination of the East European Platform. To date, in the Moldavian Platform hydrocarbons have been discovered mostly in Middle-Late Miocene (Badenian and Sarmatian) deposits, the

main fields being situated in the western part of the platform. The Badenian hydrocarbon accumulations are usually located in structural traps of faulted monocline type and the Sarmatian ones in combined traps, with a marked lithologic character due to facies variations. With the exception of Roman – Secuieni field (Sarmatian), the most important gas accumulation of the Moldavian Platform, with a discontinuous development but with a large areal extension, the other accumulations are of lesser size. In Badenian deposits, hydrocarbon accumulations are known at Cujeș, Frasin and Mălini.

The Sarmatian sands / sandstones reservoirs are exclusively gas-bearing (more than 98% methane), the most significant fields being Roman – Secuieni, Valea Seacă, Bacău and Mărgineni. In areas of the Moldavian Platform like the one considered in this study (NW part of the platform), small gas fields have been discovered through seismic surveys and exploration wells, especially during the last decade.

Thermal maturation analyses show that in the Moldavian Platform area there are two hydrocarbon

systems. The thermogenic hydrocarbon system contains source rocks of Vendian and Silurian age and oil and condensate fields hosted in the infra-anhydrite sandstone reservoirs of Badenian age located at Cujeștii, Frasin and Mălini. The biogenic hydrocarbon system is found in the Miocene formations, especially the Sarmatian ones, at depths less than 2000 m. The Upper Badenian and Sarmatian marls and shales may be considered as both source and seal rocks for this system.

The lithostratigraphic correlation of borehole data shows that the sedimentary cover of the Moldavian Platform was deposited during at least three major cycles of sedimentation (Săndulescu, 1984): (1) Late Vendian – Devonian, (2) Late Jurassic – Cretaceous – Middle Eocene, (3) Late Badenian – Sarmatian. For the scope of this study, and from the standpoint of hydrocarbon accumulations, the last sedimentation cycle is the most important one. The main lithologic character of the Badenian formations is represented by the anhydrite complex. It consists of a thick anhydrite layer which covers a complex of sands / sandstones

interlayered with shales, known as the *infra-anhydrite formation*. The Sarmatian consists of detritic formations deposited in two different sedimentary environments: deltaic and continental-lacustrine. The deltaic depositional system is characteristic for the western part of the Moldavian Platform.

During the Alpine orogeny the western part of the Moldavian Platform was gradually underthrust beneath the Eastern Carpathian Orogen. The monoclinial deposits of the Platform are dipping westward beneath the Carpathian Foredeep (molasse) and the Eastern Carpathian flysch and, also, southward (Fig. 2). The tectonic style of Moldavian Platform is dominated by a network of faults with two main directions. The first system has a NNW–SSE orientation, parallel with Eastern Carpathian orogen, and includes the most significant faults. Some of these faults affect both the basement and the sedimentary cover. The second system, mainly trending E–W or NW–SE, is younger and comprises faults of smaller displacements that affect the blocks formed by the other faults system.

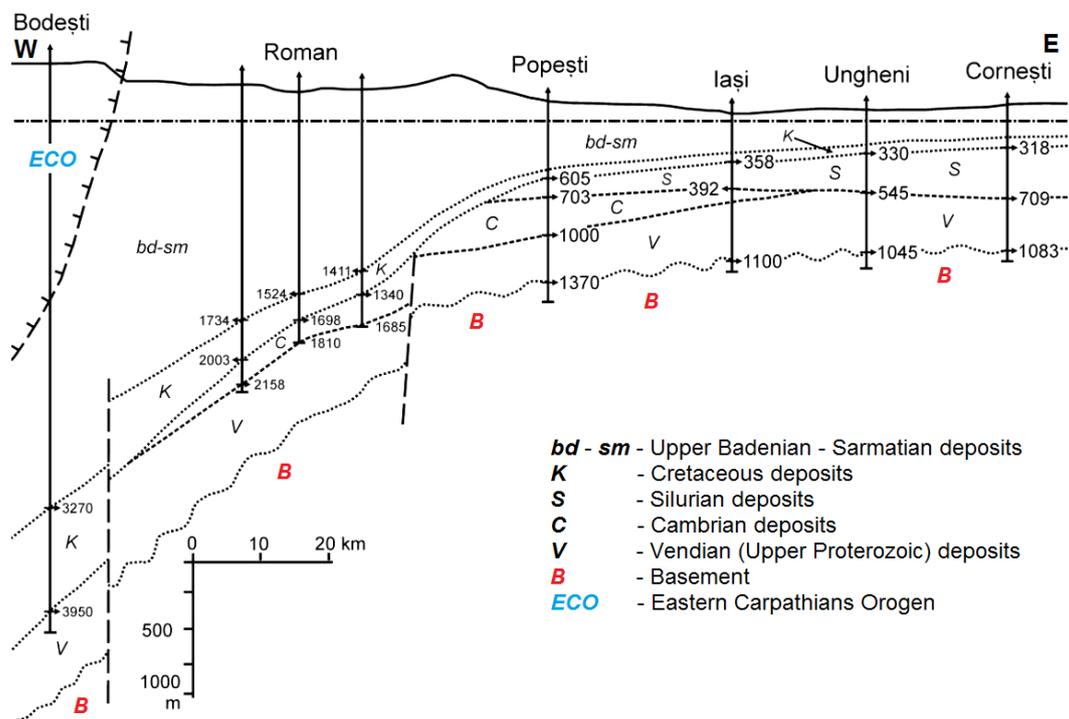


Fig. 2 – E–W cross section in the Moldavian Platform based on drilling data, showing the dip of the basement and sedimentary cover (after Pătruț and Dăneț, 1987).

The active subsidence and significant sediment supply have created favorable conditions for the accumulation of both source and reservoir rocks, as well as for the creation of conventional or subtle hydrocarbon traps.

3.2. DRILLING INFORMATION AND GEOPHYSICAL LOGGING DATA

The gas exploration well taken into consideration in this study was drilled vertically, the main exploration targets being several Sarmatian sand beds or sand bodies evidenced as sub-parallel reflectors on seismic cross sections. In the study area, the Sarmatian deposits consist of shales (calcareous and silty), siltstones, sandy siltstones and unconsolidated to partially consolidated sands/sandstones, of 5–15 m thickness. Generally, the depth of the main sand reservoirs varies between 500 m and 750 m. Secondary exploration targets for this well were represented by a Badenian sandstone section immediately underlying the Badenian anhydrite, within the infra-anhydrite formation. The Cretaceous deposits, beneath the Badenian infra-anhydrite, comprise a limestone complex (sometimes grading to calcareous sandstone), sandstones (silty to very fine, calcareous and glauconitic) which represented an additional secondary exploration target, cherts interbedded with limestone and shales.

The well was drilled in three sections with different diameters: 17.5 inch from 0 to 48 m, 12.25 inch from 48 to 305 m and 8.5 inch from 305 to 910 m (total depth). The 8.5 inch section intercepted all the exploration targets, on the stratigraphic interval Sarmatian – Cretaceous. The bottom-hole temperatures recorded in the successive wireline logging runs were 23°C at 305 m depth and 33°C at total depth. The formations tops evidenced in the *Litholog* synthetic diagram of the Mud Logging records are: 780 m – top of Badenian anhydrite, 834 m – top of Cretaceous formations.

The wireline logging program carried out in the 8.5 inch section of the borehole (drilled with KCl Polymer mud, with $\rho_m = 0.170 \text{ } \Omega\text{m @ } 20^\circ\text{C}$, $\rho_{mf} = 0.140 \text{ } \Omega\text{m @ } 20^\circ\text{C}$, $\rho_{mc} = 0.270 \text{ } \Omega\text{m @ } 20^\circ\text{C}$) consisted of: electrical logs (*SP* – spontaneous potential ΔV_{SP} [mV]; *RLLS*, *RLLD* – *Dual Laterolog* shallow and deep resistivities ρ_{RLLS} [Ωm] and ρ_{RLLD} [Ωm]; *RMLL* – *Microlaterolog* resistivity ρ_{RMLL} [Ωm]), nuclear logs (*GR* – total gamma ray intensity I_γ [API]; *NPHI* – neutron apparent porosity Φ_N [V/V]; *DEN* – bulk density δ [g/cm^3]), sonic log (*DT* – sonic compressional slowness Δt [$\mu\text{s/ft}$]) and caliper (*CAL* – borehole diameter d [in]). The geophysical logs in this section were recorded in order to determine the reservoir properties and fluid contents of the porous-permeable formations encountered in the well, to check the formation tops and to provide velocity and density data for seismic correlation.

Figure 3 presents the geophysical logs from the borehole's final section, along with a zonation track showing the *Litholog* formation tops. The Sarmatian reservoirs are delineated with respect to shales by means of low *GR* readings and positive *SP* deflections (*SP* is reversed, *i.e.* formation waters are fresher than the mud filtrate), together with a slight separation of ρ_{RLLS} and ρ_{RLLD} curves, indicating mud filtrate invasion. The Sarmatian deposits have low resistivities, ranging from 1.4 to 7.2 Ωm .

The Badenian anhydrite is clearly outlined (780–819 m depth interval) by very low *GR* values, by characteristic readings of the porosity logs ($\Phi_N \approx 0$, $\delta = 2.95\text{--}2.99 \text{ g/cm}^3$, $\Delta t = 51\text{--}56 \text{ } \mu\text{s/ft}$) and by extremely high resistivities (ρ_{RLLD} locally reaching 16000–17000 Ωm). The Cretaceous limestones complex is very well evidenced by the logs on the 834–883 m depth interval through very low *GR* values, densities reaching 2.65–2.66 g/cm^3 (together with Δt readings of 55–56 $\mu\text{s/ft}$) at the bottom, most compact, part of the complex and relatively high resistivities ($\rho_{RLLD} > 70 \text{ } \Omega\text{m}$).

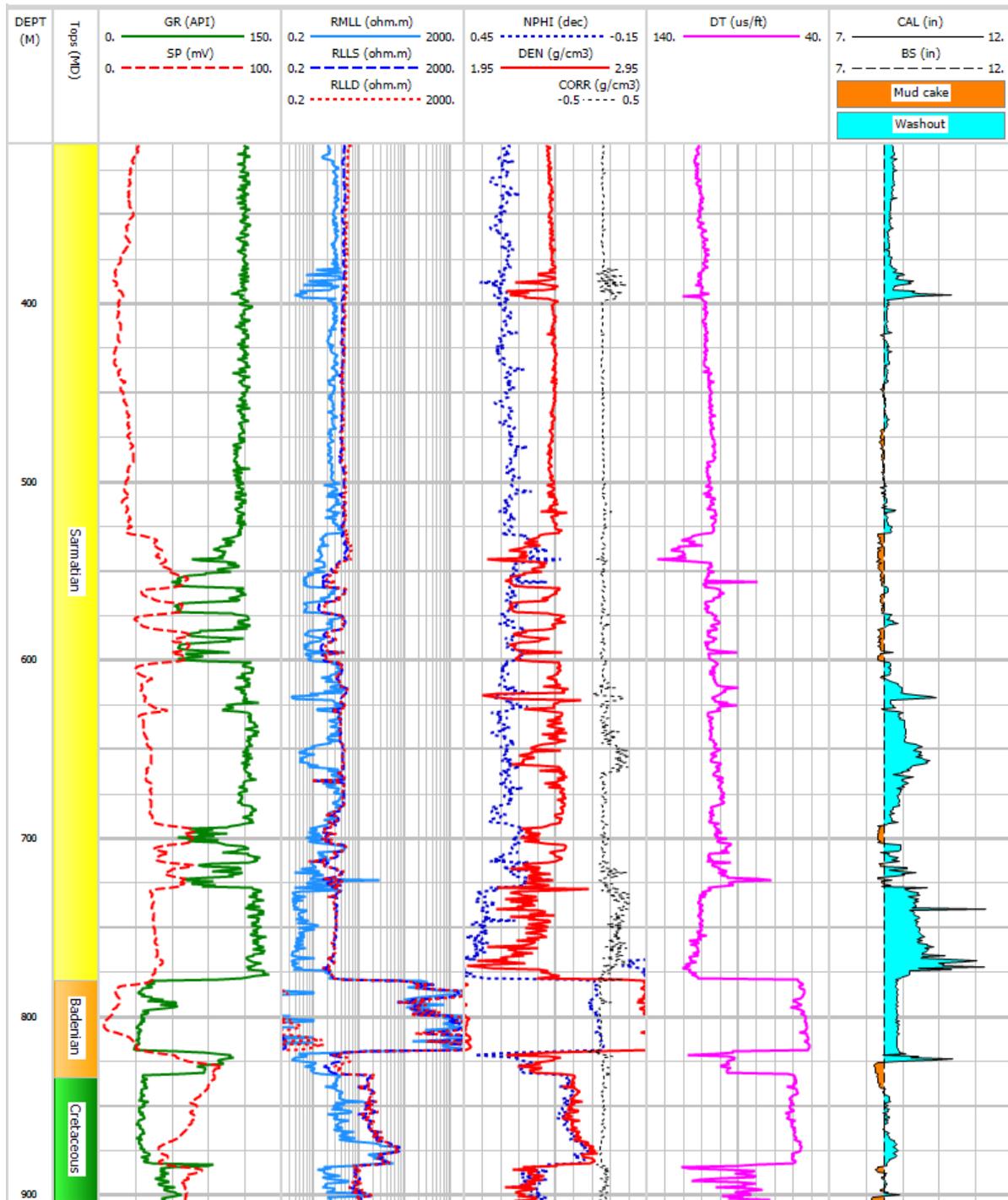


Fig. 3 – Wireline logs recorded in the analyzed well over the 8.5 inch final borehole section. Neutron porosity (NPHI) and density (DEN) logs are displayed on a standard limestone-compatible scale. The final track shows the bit size and caliper value, indicative of borehole condition.

3.3. CONVENTIONAL INTERPRETATION OF THE GEOPHYSICAL LOGGING DATA

The log interpretation challenges regarding the analyzed well consisted of:

- Complex lithology: clastics (Sarmatian), evaporites and clastics (Badenian), carbonates and clastics (Cretaceous);
- Variability of shales log responses with depth;
- Variability of formation waters resistivity (ρ_w) and salinity/salts concentration (C_w);

For the primary target, the Sarmatian deposits, initial estimates of ρ_w (and, therefore, C_w) were obtained from the amplitude of SP anomalies, in the logs pre-interpretation phase, after correcting the SP shale baseline drift with depth. The analysis was carried out for selected sand intervals (Fig. 4), assuming either predominantly NaCl formation waters or “average” fresh formation waters (for which the effect of salts other than NaCl becomes significant). Table 1 lists the results of the estimation of formation waters parameters.

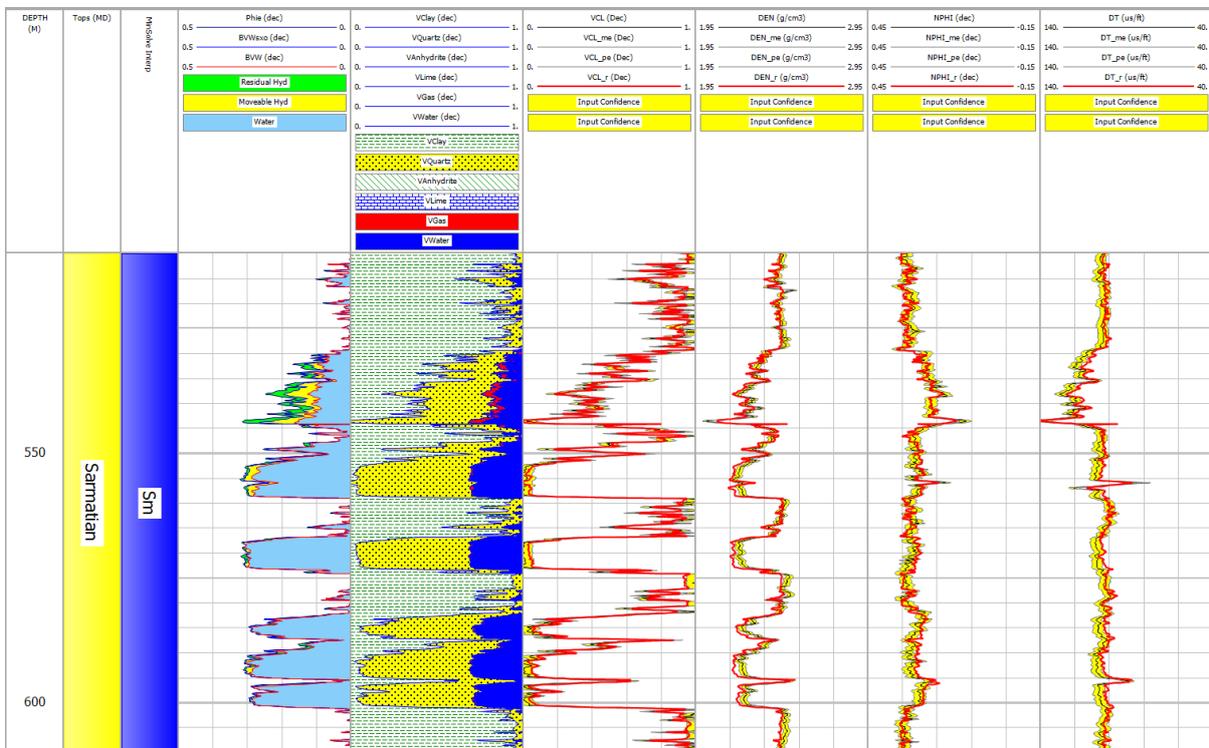


Fig. 4 – Results of the conventional interpretation of the geophysical logs on a depth interval including the main Sarmatian exploration targets. The uppermost sand is gas-bearing, the other ones below are water-bearing. The four tracks to the right show the curves/measurements used as input (in black), their reconstruction using the model's theoretical response (in red) and the uncertainty intervals assigned to each curve (yellow bands).

Table 1

Estimation of formation waters resistivity and salinity from the SP log, for selected Sarmatian sand reservoirs

Depth [m]	SP anomaly [mV]	Predominantly NaCl waters		"Average" fresh waters	
		ρ_w [Ωm]	C_w [kppm]	ρ_w [Ωm]	C_w [kppm]
553.6	+ 28	0.257	21.7	0.289	19.1
571.5	+ 25	0.230	24.2	0.255	21.7
585.7	+ 29	0.259	21.2	0.293	18.5
592.0	+ 29	0.260	21.1	0.294	18.4
598.5	+ 27	0.242	22.7	0.271	20.1

In the bottom part of the 8.5 inch borehole section the SP curve is almost featureless, with typical highly resistive formations signature (linear variation in the compact Badenian anhydrite and the Cretaceous limestone); this makes the SP log unusable for ρ_w and C_w estimation. The lack of separation for ρ_{LLS} and ρ_{LLD} curves most likely indicates deep invasion in low-porosity intervals. Also, the Φ_N and δ curves are superimposed on a limestone-compatible scale, showing no obvious hydrocarbon effects (neutron-density crossover) and probably indicating water-bearing rocks. Overall, there are no “quick look” hydrocarbon indications in this borehole interval. The formation waters resistivity for this interval was estimated during the interpretation, from a $\log(\Phi) = f(\log(\rho_{LLD}))$ Pickett crossplot using the computed effective porosity Φ . Multiple ρ_w trends resulted from the porosity – resistivity crossplot for the Badenian and Cretaceous formations.

The ρ_w values finally used in the interpretation range from 0.29 Ωm (Sarmatian) to 0.55 Ωm (Cretaceous). In addition, the best interpretation

results were obtained by using multiple values for the cementation exponent m (ranging from 1.5 to 2.0) in the Sarmatian, Badenian and Cretaceous formations, instead of a single m . The rest of Archie's parameters, *i.e.* tortuosity factor a and saturation exponent n , were set to 1.0 and, respectively, 2.0.

For the interpretation, the final borehole section was divided into five zones (Fig. 4 and Fig. 5): Sm – Sarmatian, Bd_1 – Badenian anhydrite, Bd_2 – Badenian infra-anhydrite formation, K_1 – Cretaceous limestone complex, K_2 – lower Cretaceous sandstones. The interpretation was carried out using the probabilistic module “*Mineral Solver*” included in *Interactive Petrophysics (IPTM)* software (© LR Senergy). The module solves the system of equations representing the responses of logging tools with respect to a certain petrophysical model comprising solid and fluid volume fractions. The solution (mineralogy, porosity, fluid saturations) obtained at each depth level is the most probable, *i.e.* optimal.

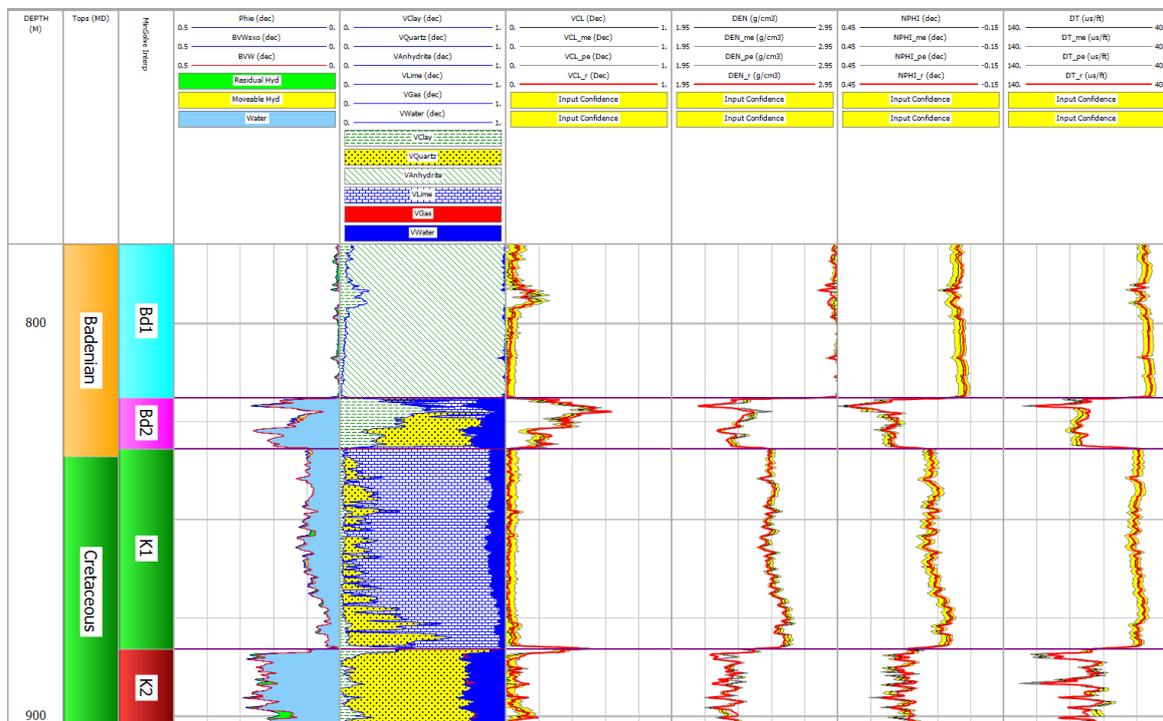


Fig. 5 – Results of the conventional interpretation of the geophysical logs on a depth interval including the secondary exploration targets in the Badenian and Cretaceous formations. The porous-permeable formations intercepted by the well on this interval are water-bearing.

A variable uncertainty (acting as a weighting factor) is assigned to each logging tool, to take into consideration the relative importance of one response equation to another and, also, to mitigate the effect of bad hole intervals. The response equations end-points (100% minerals/fluids readings) for certain components, such as clay, clean matrix, formation water parameters or hydrocarbons parameters, are set based on logs pre-interpretation.

The interpretation's quality and accuracy are evaluated by comparing the reconstructed tool responses (synthetic logs) to the original input tool responses (measured logs), using a global error function. The adjustment of the end-point parameters and/or the interpretation model (number and type of solid and fluid volume fractions) allow the best possible log input data reconstruction at each depth level.

For computing the water saturations in the uninvaded and the flushed zone of porous-permeable formations, the "Indonesia" (Poupon and Leveaux, 1971) equation for shaly formations was used. The clay volume (V_{cl}) was estimated from a combination of clay indicators (GR and the $\delta = f(\Phi_N)$ crossplot) and the clay resistivity, seen by the deep and the very shallow investigation tools, was estimated from $V_{cl} = f(\rho_{LLD})$ and $V_{cl} = f(\rho_{MLL})$ crossplots.

The log interpretation results for the 8.5 inch borehole section are presented in Fig. 4 and Fig. 5. Gas was identified only in the uppermost Sarmatian sand reservoir (530–545 m depth interval). A flow test carried out for this reservoir confirmed the interpretation, producing dry gas at commercial rates.

3.4. PRINCIPAL COMPONENT ANALYSIS OF THE GEOPHYSICAL LOGGING DATA

The PCA was carried out on the same depth interval as the conventional log interpretation (305–910 m, the 8.5 inch borehole section), in order to compare the results.

PCA can be performed using the covariance matrix Σ of the data set or, alternately, using the correlation matrix R . If the data (the geophysical logs) are normalized by removing the mean values μ and taking as unity the standard deviations σ , the covariance matrix becomes the correlation matrix. As an example, for two logs (data vectors) x and y with N samples, mean values μ_x , μ_y and standard deviations σ_x , σ_y , the correlation coefficient r is defined by:

$$r(x, y) = \frac{\text{cov}(x, y)}{\sigma_x \sigma_y} = \frac{\sum_{i=1}^N [(x_i - \mu_x)(y_i - \mu_y)]}{\sqrt{\sum_{i=1}^N (x_i - \mu_x)^2} \sqrt{\sum_{i=1}^N (y_i - \mu_y)^2}} \quad (5)$$

Table 2 lists the elements of the covariance / correlation matrix of the entire data set (excluding the SP and CAL logs, which are not suitable for a principal component analysis).

The correlation coefficient values in Table 2 may be evaluated using the following criteria: very high correlation: $r = 0.9-1.0$; high correlation: $r = 0.7-0.9$; moderate correlation: $r = 0.5-0.7$; low correlation: $r = 0.3-0.5$; little or no correlation: $r = 0.0-0.3$.

The logs effectively used as input for PCA were GR , $RMLL$, $RLLD$, $NPHI$, DEN and DT (6 logs with 6050 data samples/log). The PCA results are presented in Table 3 and a comparison between the results of conventional log interpretation and the PCA results is presented in Fig. 6 (the score logs z_i of the principal components are expressed in standard deviation units).

Table 2

The covariance/correlation matrix of the complete geophysical logs data set (7 logs with 6050 data samples/log)

	<i>GR</i>	<i>RLLD</i>	<i>RLLS</i>	<i>RMLL</i>	<i>NPHI</i>	<i>DEN</i>	<i>DT</i>
<i>GR</i>	1	-0.6801	-0.6860	-0.5670	0.8877	-0.4956	0.8645
<i>RLLD</i>		1	0.9939	0.9137	-0.8285	0.8405	-0.7956
<i>RLLS</i>			1	0.9225	-0.8391	0.8570	-0.8184
<i>RMLL</i>				1	-0.7691	0.8948	-0.7225
<i>NPHI</i>					1	-0.7249	0.9016
<i>DEN</i>						1	-0.7384
<i>DT</i>							1

Table 3

Principal components of the geophysical logs covariance/correlation matrix

	Variances explained by principal components (eigenvalues) [% of total data variance]					
	PC ₁	PC ₂	PC ₃	PC ₄	PC ₅	PC ₆
	81.69	12.26	2.81	1.54	1.01	0.69
	Component loadings (eigenvectors)					
	PC ₁	PC ₂	PC ₃	PC ₄	PC ₅	PC ₆
GR	0.37268	0.62829	-0.22511	0.06773	-0.38833	-0.51019
RLLD	-0.42330	0.22487	0.51642	0.54463	-0.43241	0.14128
RMLL	-0.40748	0.43385	0.30454	-0.20671	0.60248	-0.38377
NPHI	0.42738	0.25964	-0.03796	0.60030	0.53286	0.32279
DEN	-0.39694	0.47338	-0.52716	-0.19476	-0.06154	0.54657
DT	0.41912	0.27380	0.55727	-0.50772	-0.10731	0.41173

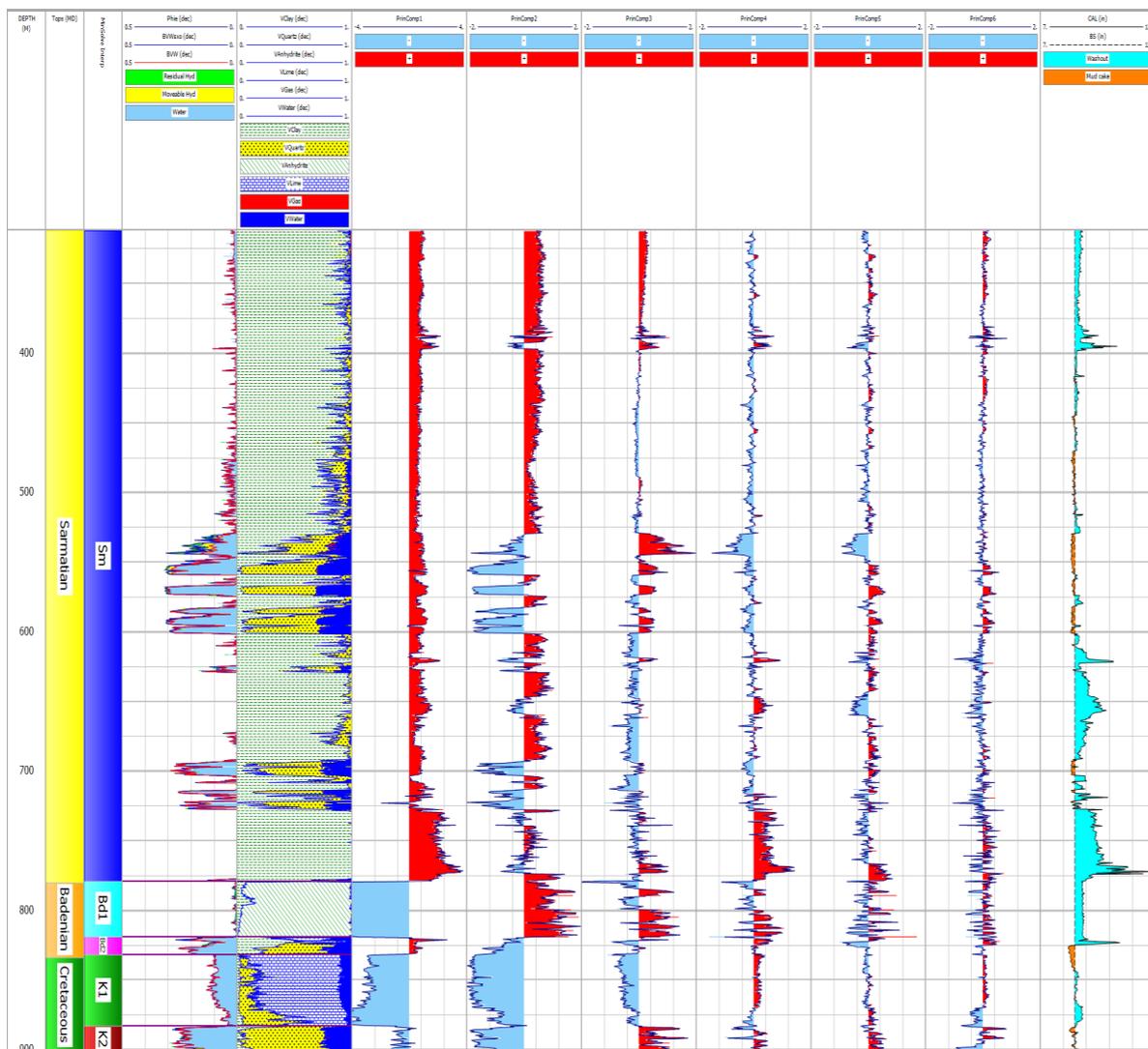


Fig. 6 – Comparison between the results of conventional log interpretation (track 4 – effective porosity and bulk volumes of fluids, track 5 – volumetric formation analysis) and the PCA results (tracks 6 to 11 – score logs of the principal components PC₁ to PC₆). Track 2 shows the actual formation tops and track 3 shows the zoning used for interpretation. The caliper log and bit size are presented in track 12, to illustrate the borehole condition.

The first principal component of the geophysical logs (PC_1) explains the largest part (81.69%) of the total variability in the data set, with approximately equal loadings (weights) for all logs. $RLLD$, $RMLL$ and DEN (negative weights) are inversely correlated with GR , $NPFI$ and DT (positive weights). In track 6 from Fig. 6, the depth intervals with positive PC_1 score log correspond to formations with high GR , $NPFI$ and DT , but low $RLLD$, $RMLL$ and DEN , while the negative PC_1 scores delineate the opposite case. PC_1 acts as a major lithological “cut-off”, separating the younger and/or less compact formations (Sarmatian and Badenian shales, sands and slightly cemented sandstones) from the older and/or compact, low-porosity and resistive formations (Badenian anhydrites, Cretaceous limestones and highly cemented sandstones).

The second principal component (PC_2 – Fig. 6, track 7) accounts for 12.26% of the total variability in the log suite, being dominated by the contribution of GR and subordinately DEN . PC_2 may be interpreted as an accurate separator of porous-permeable intervals (negative score values), no matter their lithological composition, with respect to impermeable formations (positive score values) – shales or very compact rocks. The reservoir boundaries are accurately delineated by the strong and sudden sign variations/changes of PC_2 synthetic score log. With the reservoirs once separated by PC_2 , all further PCs interpretations in terms of additional petrophysical information (e.g., fluids identification) should be focused only on these zones.

Higher-order components, like PC_3 to PC_5 , respond more to fluids type and volume (or reflect other fluid-related influences), as a result of significant $RLLD$ and $RMLL$ loadings in their eigenvectors structure. PC_3 (Fig. 6, track 8) explains 2.81% of total data variability unrelated to PC_1 and PC_2 . It has higher loadings for $RLLD$, DEN and DT , DEN being inversely correlated with $RLLD$ and DT ; the $RLLD$ contribution indicates a fluid-related response (type and/or

volume). The positive PC_3 score values correspond to formations showing relatively high resistivity, low bulk density and high sonic transit time, *i.e.* good indicators of hydrocarbons presence (particularly light hydrocarbons, such as gas). The strong positive PC_3 “anomaly” noticeable in the 529–545 m interval (Sarmatian gas-bearing sand) correlates extremely well with the conventional log interpretation results and flow test results.

The components PC_4 and PC_5 (Fig. 6, tracks 9 and 10) explain 1.54% and, respectively, 1.01% of total data variance, unrelated to PC_1 , ..., PC_3 . The $RLLD$, $RMLL$, $NPFI$ and DT logs have important contributions in the eigenvectors structure; most likely, PC_4 and PC_5 respond to intermediate and shallow-depth (flushed zone) fluid-related factors (fluids type and/or volume). Significant PC_4 and PC_5 negative score “anomalies” are seen on the 529–545 m interval (the uppermost Sarmatian gas-bearing sand). The PC_5 score “anomaly” is negative only in the gas-bearing sand and positive in all other Sarmatian sands, which are water-bearing. Presumably, the PC_4 and PC_5 negative “anomalies” are related to the large neutron log contribution in the corresponding eigenvectors and to the low hydrogen index (neutron porosity) of gas with respect to formation water. In the Badenian and Cretaceous reservoirs PC_4 and PC_5 score “anomalies” have opposite sign (positive) compared to the “anomalies” in the Sarmatian gas-bearing sand (negative) and they have very low amplitudes.

Figure 7 shows a detailed comparison between the conventional log interpretation results and the PCA results for a 100 m depth interval including the upper Sarmatian sands. This allows a clear evaluation of the characteristic “anomalies” which appear on the synthetic score logs z_i for each reservoir, illustrating both the PCA capacity of separating them with respect to the impermeable formations (shales), as well as the possibility of fluid type assessment, by comparing the “anomalies” obtained for several reservoirs.

REFERENCES

- BARRASH, W., MORIN, R.H. (1997), *Recognition of units in coarse, unconsolidated braided-stream deposits from geophysical log data with principal components analysis*. *Geology*, **25** (8), 687–690.
- GONÇALVES, C.A. (1998), *Lithologic Interpretation of Downhole Logging Data from the Côte d'Ivoire – Ghana Transform Margin: A Statistical Approach*. In: Mascle, J., Lohmann, G.P. and Moullade, M. (eds.), *Proceedings of the Ocean Drilling Program, Scientific Results*, **159**, 157–170. Ocean Drilling Program, College Station, TX, USA.
- HOTELLING, H. (1933), *Analysis of a complex of statistical variables into principal components*. *Journal of Educational Psychology*, **24** (6), 417–441.
- JOLLIFFE, I.T. (2002), *Principal Component Analysis*, Second Edition, Springer Series in Statistics, Springer-Verlag.
- KASSENAAR, J.D.C. (1991), *An application of principal components analysis to borehole geophysical data*. 4th International MGLS/KEGS Symposium on Borehole Geophysics for Minerals, Geotechnical and Groundwater Applications, Proceedings.
- LIM, J.S., KANG, J.M., KIM, J. (1998), *Artificial Intelligence Approach for Well-to-Well Log Correlation*. SPE India Oil and Gas Conference and Exhibition, SPE-39541-MS.
- MOLINE, G.R., BAHR, J.M., DRZEWEJECKI, P.A., SHEPHERD, L.D. (1992), *Identification and characterization of pressure seals through the use of wireline logs: A multivariate statistical approach*. *The Log Analyst*, **34**, 362–372.
- MORIN, R.H. (2006), *Negative correlation between porosity and hydraulic conductivity in sand-and-gravel aquifers at Cape Cod, Massachusetts, USA*. *Journal of Hydrology*, **316**, 43–52.
- PĂTRUȚ, I., DANEȚ, TH. (1987), *Le Pre-cambrien (Vendien) et le Cambrien dans la Plate-forme Moldave*. *Scientific Annals of the “Alexandru Ioan Cuza” University, Iași, Romania, Section II–b. Geology-Geography*, Tome XXXIII, 26–30.
- PEARSON, K. (1901), *On Lines and Planes of Closest Fit to Systems of Points in Space*. *Philosophical Magazine*, **2**, 559–572.
- POUPON, A., LEVEAUX, J. (1971), *Evaluation of Water Saturation in Shaly Formations*. *The Log Analyst*, **12** (4), 3–8.
- SÂNDULESCU, M. (1984), *Geotectonics of Romania*. Technical Publishing House, Bucharest, Romania. (in Romanian).

Received: February 7, 2018

Accepted for publication: March 5, 2018